

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penyakit liver merupakan penyakit peradangan pada organ hati, secara umum faktor penyebab terjadinya penyakit liver dapat disebabkan oleh pola hidup yang tidak sehat namun faktor lainnya adalah kondisi adanya kelainan hati yang merupakan bawaan sejak lahir atau pada saat kelahiran, adanya gangguan dan kelainan pada proses metabolisme, terinfeksi virus atau bakteri, kekurangan gizi atau nutrisi, ketergantungan alkohol dan zat adiktif lainnya maupun kecanduan dan kebiasaan merokok juga dapat menjadi penyebab dari penyakit liver (Radhakrishnan, Sridevi, dan D, 2015).

Permasalahan yang dihadapi masyarakat saat ini salah satunya adalah keterlambatan penanganan secara medis kepada penderita liver karena sebagian besar pasien memeriksakan kondisinya setelah penyakit terdeteksi sudah pada stadium lanjut. Untuk menanggulangi masalah terjadi semakin parahnya kondisi kesehatan penderita maka diperlukan pemeriksaan rutin dan pencegahan resiko adanya serangan penyakit kronis tersebut, namun hal tersebut tidak dilakukan oleh sebagian masyarakat karena beberapa alasan diantaranya rutinitas yang padat, mahalnya biaya pemeriksaan serta takut akan adanya diagnosa penyakit kronis.

Data yang tidak konsisten pada rekam medik di karenakan oleh banyaknya data yang hilang, akan dapat membatasi analisis dan interpretabilitas sistem terkait dengan dampak *relative* dari data spesifik elemen (Li, Hutfless, Scharfstein, Daniels, Hogan, Little, 2014). Identifikasi jenis data yang hilang sangat penting untuk menemukan solusi untuk mengatasinya. Data yang hilang telah dikategorikan menjadi tiga berbagai jenis yaitu: hilang sepenuhnya secara acak (MCAR), hilang secara acak (MAR) dan hilang tidak secara acak (MNAR) (Little, Rubin, 2000). MCAR didefinisikan sebagai data yang ketidakhadirannya terjadi terlepas dari yang diamati dan / atau nilai yang tidak teramati (Sterne, White, Carlin, Spratt, Royston, Kenward, 2009). Misalnya, berat badan seorang individu tidak dicatat karena skala yang digunakan untuk menangkap berat dinonaktifkan sementara. Data yang hilang mengurangi ukuran sampel populasi penelitian yang tersedia dan kemudian mengurangi statistik kekuatan, meskipun tanpa adanya bias (Mack, Su, 2018). Sebaliknya, MAR terjadi ketika alasan untuk titik data yang hilang dapat disimpulkan oleh hanya nilai yang diamati dan tidak terkait dengan nilai yang tidak teramati.

Menghapus baris data karena tidak adanya beberapa variabel menyebabkan kerugian informasi berharga yang diamati yang akan bermanfaat bagi analisis. Ini dapat menyebabkan estimasi yang bias. Data yang hilang juga dapat dilengkapi dengan temuan dari catatan klinis menggunakan teknik Natural Language Processing (NLP) (Hegde, Shimpi, Glurich, Acharya, 2018). Atau, mencoba untuk memprediksi nilai yang hilang dengan estimasi dekat berdasarkan konteks data disebut data imputasi (Newgard, Lewis, 2015). Teknik telah dikembangkan untuk menghubungkan elemen data yang hilang. Ini menggunakan analisis statistik pada informasi yang diketahui bersama dengan penanganan ketidakpastian yang disebabkan oleh adanya data yang hilang untuk menghasilkan perkiraan (Manly, Wells, 2015).

MICE dapat digunakan untuk berbagai model data, seperti data kontinu, data biner (regresi logistik), data kontinu 2-level, regresi logistik polikotomus, dan odds proporsional (Buuren, Groothuis, 2011). Prosedur MICE mengikuti serangkaian model regresi yang dijalankan, dimana masing-masing variabel dari data yang hilang dimodelkan bersyarat pada variabel lain dalam data tersebut.

1.2 Rumusan Masalah

Rumusan masalah yang ada dapat di definisikan sebagai berikut:

1. Bagaimana caranya memfilter data liver yang kotor agar tidak tercampur dengan data penyakit yang lain?
2. Bagaimana caranya memperbaiki *missing* data pada data rekam medik liver yang sudah terfilter?

1.3 Tujuan Penelitian

Adapun tujuan dari skripsi ini adalah untuk :

1. Membersihkan data liver yang kotor atau yang tercampur dengan data yang lain untuk mendapat hasil data yang bisa digunakan.
2. Menerapkan algoritma MICE, MEAN dan MEDIAN untuk memperbaiki missing data.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari skripsi ini adalah :

1. Menambah variasi pembelajaran dalam proses mengelolah data (MICE MEAN dan MEDIAN).
2. Memudahkan kita dalam memfilter data.
3. Hasil penelitian ini diharapkan dapat bermanfaat sebagai pembelajaran mengelolah data di bidang Informatika.